

Visual Alignment

3D Residual Formulas; Silicon Microtracker Examples

Thomas G. Trippe

Introduction

The process of aligning detectors using reconstructed tracks is complex. It requires careful selection of the events, track samples, and geometry elements to be used in each step of the alignment. It requires a well designed geometry system, such as the DØ Geometry System, which allows control over the geometry parameters that enter the reconstruction. It requires the basic formulas for computing the changes in the geometry parameters needed to align the selected geometry elements and a means of evaluating the improvement in the alignment due to these changes. It requires development of a strategy for the sequence of successive alignments needed to produce a fully aligned detector.

This note focuses on the basic alignment formulas and on techniques for visualizing residuals. To do this, we introduce the concept of 3D residual vectors that can be displayed to reveal misalignments, as well as used in calculations of alignment parameters.

In track fitting, a residual is the distance between a track's measured position and its fitted position. Such a residual can be represented as a 3-vector from the measurement to the track along the closest distance of approach between them. The error matrix for this residual includes information on the precision of the measurement and of the fitted track position, including the large (in principle infinite) uncertainty along the track direction. For example, a single-sided silicon strip measurement is a line and the fitted track is a line. The 3D residual vector represents the closest distance of approach between these two lines. This 3D residual does not usually lie in the plane of the silicon detector, in contrast to the usual 1D residual, which is usually defined as the signed scalar distance from the track to the silicon strip in the plane of the detector. Note that the sign convention chosen here has 3D residual vectors pointing from measurements to fitted tracks, so that the 3D residuals and their positions on the detector indicate the magnitude and direction of the movements needed to align the detector.

The advantage of 3D residuals with error matrices is that they contain all of the information about the track orientation and the detector geometry needed for alignment. In the 1D viewpoint, that information resides in the track orientation and detector geometry instead of in the residuals themselves, so that to calculate χ^2 , one must use the track orientation and detector geometry in addition to the residuals. The 1D residuals alone cannot give any indications of misalignments because they contain no direction information.

The 3D residual viewpoint seems well adapted for visualization of alignment problems. We can display a field of 3D residual vectors and their error ellipses more easily than displaying 1D residuals, tracks, and detector measurements. Since the 3D residual vectors give the displacements at various points of the detector needed to align it, combined with a random component, we can look at this residual field to decide whether the detector is significantly out of alignment, and if so, which parts of the detector need alignment.

Unless the misalignment is large compared with the measurement errors, it may not be possible to get significant visual information from the individual residuals. However, we can combine the effects of the residuals on a given part of the detector by doing a χ^2 minimization to determine the translation and rotation needed for alignment. We can then display this translation and rotation for each detector part, instead of displaying all of the individual residuals. The translation and rotation represent averages over the residual field. For example, in the absence of rotations, the resulting translation is just the weighted average of the residuals.

If it proves to be useful, we could even develop an interactive alignment program in which we could move the detector components by the amount indicated by the 3D residuals and would get back a set of alignment parameters and an improved set of 3D residuals. This could be of use during the initial zeroth order alignment.

A major advantage of the 3D residual viewpoint is that it is fairly detector independent. The 1D residuals are detector specific, in the sense that they must be used in conjunction with track orientations and with detector geometry (e.g. strip orientations) to determine alignment parameters. In contrast, the 3D residual fields are detector independent, in the sense that they contain all information about track angles and detector geometry that are needed to determine alignment parameters. The concepts discussed here for 3D residuals and their error matrices are detector independent. The alignment formulas are detector independent, as long as the detector elements are rigid objects that can be aligned by rotation and translation. This note gives the formulas for calculating alignment parameters from 3D residuals, assuming that the detector elements are rigid objects. Examples are given for silicon detectors.

The formulas for combining residuals to determine non-rigid transformations such as scaling, distortion, and changing strip pitch are detector dependent and need to be developed separately for each case.

The 3D residuals themselves are not entirely detector independent. Each residual is associated with a detector element. That allows us to select the residuals associated with a particular part of the detector. It also gives us access to the local coordinate system of that part of the detector so that we can constrain the alignment parameters, for example to allow motion only in the plane of a silicon ladder.

Note that in defining 3D residuals, the use of the 3-vector closest distance of approach between the measurement and the track is not essential. We could equally well have chosen the 3-vector between the measurement and the track in the plane of the detector, as long as we retain the correct three-dimensional error. The error ellipse for this error matrix has infinite extent along the track direction, so that the χ^2 contribution of a given residual is completely insensitive to whether the residual touches the track at the closest distance of approach or in the plane of the detector. The closest distance of approach seems preferable from a visualization standpoint, although some might prefer working in the plane of the detector, since it is more typical to evaluate residuals in this plane.

Formulas for 3D residual vectors

Let \mathbf{d} be a residual vector pointing from track measurement position \mathbf{s} to fitted track position \mathbf{t} , chosen so that $\mathbf{d} = \mathbf{t} - \mathbf{s}$ and $|\mathbf{d}|$ is the closest distance of approach between the measurement and the fitted track. Uncertainties in the measurement and fitted track lead to uncertainties in \mathbf{d} . The uncertainties in the three components of \mathbf{d} are given by the covariance matrix, or error matrix, \mathbf{E} , whose diagonal elements are the squares of the uncertainties in d_x , d_y , and d_z , and whose off-diagonal elements are the correlated uncertainties, which can be significant, especially when they represent the very large uncertainties along the track direction. For a collection of residual vectors \mathbf{d}_i and their covariance matrices \mathbf{E}_i , we can calculate the χ^2 that these residuals are statistical fluctuations:

$$\chi^2 = \sum_i \mathbf{d}_i^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i ,$$

$$\mathbf{d}_i = \mathbf{t}_i - \mathbf{s}_i .$$

where i runs over all measurements of all tracks being considered in a particular alignment step. Matrices and vectors are written as bold-faced upper and lower case letters, respectively. In products of vectors and matrices, summations over the coordinate indices are implied. This χ^2 can be minimized with respect to alignment parameters, such as the translations and rotations of components of the detector, to obtain best values of the alignment parameters.

If the detector part is rotated by \mathbf{R} and then translated by \mathbf{h} , Appendix A shows that the residuals \mathbf{d}_i are transformed into \mathbf{d}'_i according to

$$\mathbf{d}'_i = \mathbf{d}_i - \mathbf{Q}\mathbf{s}_i - \mathbf{h},$$

$$\mathbf{Q} = \mathbf{R} - \mathbf{1}.$$

The $\chi^2(\mathbf{Q}, \mathbf{h})$ that the \mathbf{d}'_i , the residuals after rotation \mathbf{R} and translation \mathbf{h} , are equal to zero within their error matrices \mathbf{E}_i is

$$\chi^2(\mathbf{Q}, \mathbf{h}) = \sum_i \mathbf{d}'_i{}^\dagger \mathbf{E}_i^{-1} \mathbf{d}'_i.$$

The \mathbf{d}'_i are linear in the translation components and, for small angles, the rotation angles, so that $\chi^2(\mathbf{Q}, \mathbf{h})$ can be minimized analytically. We linearize \mathbf{Q} for small rotations θ_j around the x, y, and z axes by writing $\mathbf{Q} = \theta_j \mathbf{Q}_j$, with

$$\mathbf{Q}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \mathbf{Q}_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Differentiating $\chi^2(\mathbf{Q}, \mathbf{h})$ with respect to the three rotation angles θ_j of \mathbf{Q} and the three components h_k of \mathbf{h} and setting the differentials equal to zero yields six equations in the six parameters θ_m and h_n :

$$\sum_{m=1}^3 A_{jm} \theta_m + \sum_{n=1}^3 B_{jn} h_n = c_j, \quad j=1,3,$$

$$\sum_{m=1}^3 D_{jm} \theta_m + \sum_{n=1}^3 F_{jn} h_n = g_j, \quad j=1,3,$$

$$A_{jm} = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_m \mathbf{s}_i, \quad B_{jn} = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_n, \quad c_j = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i,$$

$$D_{jm} = \sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_m \mathbf{s}_i, \quad F_{jn} = \sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_n, \quad g_j = \sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i.$$

These six equations can be written as a single six-dimensional matrix equation

$$\mathbf{M}\mathbf{p} = \mathbf{k}$$

where the six-by-six matrix \mathbf{M} and six-vectors \mathbf{p} of parameters and \mathbf{k} of constants are given by

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{D} & \mathbf{F} \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{h} \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} \mathbf{c} \\ \mathbf{g} \end{bmatrix}.$$

The minimum χ^2 solution for \mathbf{p} , the rotation-translation six-vector, is

$$\mathbf{p} = \mathbf{M}^{-1} \mathbf{k}.$$

Appendix A shows that $\mathbf{M}^{-1} = \mathbf{V}$, the error matrix of the parameters \mathbf{p} , so that the solution can be written

$$\mathbf{p} = \mathbf{V} \mathbf{k}$$

which expresses the parameters \mathbf{p} of the χ^2 minimum in terms of their error matrix \mathbf{V} times the constant vector \mathbf{k} .

Special case: pure translation, no rotation

To understand these formulas, consider some special cases. The results are shown here but more details are given in Appendix A. If we set the rotation \mathbf{Q} to zero, then only F_{jm} and g_j are non-zero, yielding

$$\mathbf{h} = \left(\sum_i \mathbf{E}_i^{-1} \right)^{-1} \sum_i \mathbf{E}_i^{-1} \mathbf{d}_i,$$

In other words, for purely translational alignment, \mathbf{h} is the weighted average of the residuals, weighted by the inverses of their error matrices. The error matrix \mathbf{E} for \mathbf{h} is just the inverse of the sum of the weights.

$$\mathbf{E} = \left(\sum_i \mathbf{E}_i^{-1} \right)^{-1}.$$

Special case: pure rotation, no translation

Another special case is that of purely rotational alignment, with translation \mathbf{h} zero, so that only A_{jm} and c_j enter, yielding

$$\sum_{m=1}^3 A_{jm} \theta_m = c_j, \quad j=1,3,$$

$$A_{jm} = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_m \mathbf{s}_i, \quad c_j = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i$$

$$\theta_m = \left(\sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_m \mathbf{s}_i \right)^{-1} \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i$$

or, in matrix-vector notation,

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{c},$$

$$\boldsymbol{\theta} = \mathbf{A}^{-1}\mathbf{c}.$$

The rotation vector $\boldsymbol{\theta}$ is, like for translation \mathbf{h} above, a weighted average over the residuals \mathbf{d}_i , but with a more complex weighting, discussed in more detail in Appendix A.

Special case: translation in one dimension only

Another special case is that of allowing only a single parameter to vary. If only h_j , the j^{th} component of \mathbf{h} varies, then only a single equation, $\frac{\partial \chi^2}{\partial h_j} = 0$, replaces the six equations derived earlier, giving

$$\sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} (\mathbf{d}_i - h_j \mathbf{u}_j) = 0$$

or

$$h_j = \left(\sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_j \right)^{-1} \sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i$$

The inverse error squared in h_j is given by

$$\frac{1}{\sigma_{h_j}^2} = \frac{1}{2} \left(\frac{\partial^2 \chi^2}{\partial h_j \partial h_j} \right) = \sum_i (\mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_j)$$

Special case: rotation in one dimension only

Likewise, if we fix all but a single angular parameter θ_j are fixed, then only a single equation $\frac{\partial \chi^2}{\partial \theta_j} = 0$ remains, giving

$$\theta_j = \left(\sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_j \mathbf{s}_i \right)^{-1} \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i$$

with inverse error squared in θ_j given by

$$\frac{1}{\sigma_{\theta_j}^2} = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_j \mathbf{s}_i$$

Histogramming residuals

As a check on the residuals it is useful to histogram them to see their distributions in one dimension. The distributions of residuals, normalized by their errors should be Gaussian and peaked at zero if there are no misalignments. If there are misalignments, the distribution of unnormalized residuals should peak at a point whose distance from zero reveals the magnitude of the misalignment. The unnormalized residuals corresponding to the three translation parameters h_j are

$$h_j^i = (\mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_j)^{-1} \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i, \quad j=1,3, i=1, N_{\text{residuals}}$$

The inverse error squared in h_j is given by

$$\frac{1}{\sigma_{h_j^i}^2} = \sum_i (\mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_j), \quad j=1,3, i=1, N_{\text{residuals}}$$

These are just the components of the weighted averages of residuals in one dimension given above, so that

$$h_j = \left(\sum_i \frac{1}{\sigma_{h_j^i}^2} \right)^{-1} \sum_i \frac{1}{\sigma_{h_j^i}^2} h_j^i, \quad j=1,3, i=1, N_{\text{residuals}}$$

Similarly, the unnormalized residuals corresponding to the three rotation angles θ_j are

$$\theta_j^i = (\mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_j \mathbf{s}_i)^{-1} \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i, \quad j=1,3, i=1, N_{\text{residuals}}$$

with inverse error squared in θ_j given by

$$\frac{1}{\sigma_{\theta_j^i}^2} = \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_j \mathbf{s}_i, \quad j=1,3, i=1, N_{\text{residuals}}.$$

Example of Residual Formulas- Silicon Detector

A track passing through a silicon wafer is expected to give rise to signals on a cluster of one or more strips, or for double-sided silicon, a cluster on either side. The clusters from double-sided silicon are joined into 3D hits for DØ track finding. In the following example, we do not join the clusters, but treat each side as a separate hit with a huge error along the strip direction. This allows a simpler illustration of the power of the 3D residual error matrices to correctly contain strip (and track) direction information

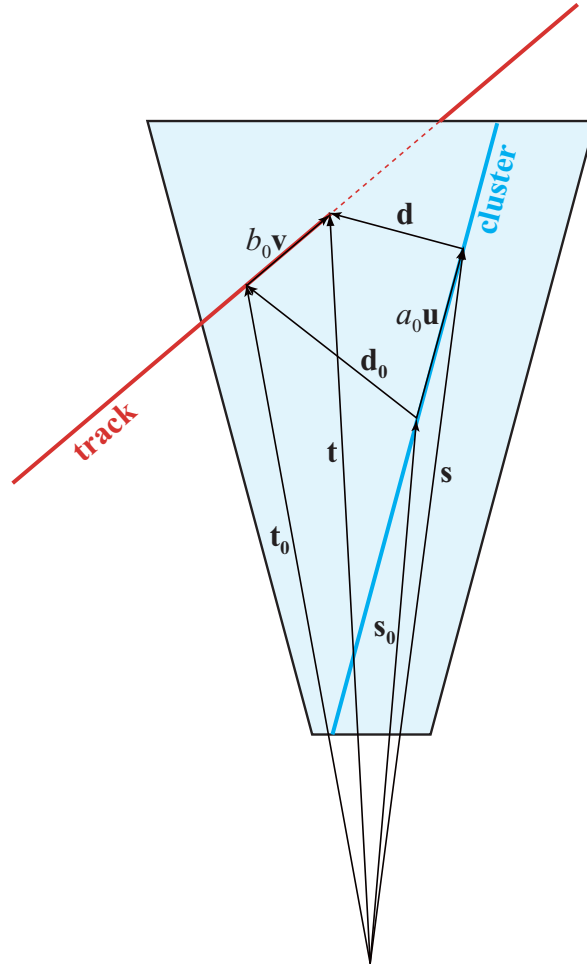


Figure 1. A single-sided silicon wedge, a track normal to it and a cluster displaced from the track by the 3D residual vector \mathbf{d} . See text for details.

A fitted track will probably not pass directly through its cluster(s), due to reconstruction errors and possibly to misalignment. Figure 1 shows a single-sided silicon wedge with a track significantly displaced from the cluster due to misalignment.

The three-vector distance \mathbf{d} from the cluster to the fitted track is the 3D residual, and can be derived as follows. Any point $s(a)$ on a cluster can be represented by a vector $\mathbf{s}(a)$ from the origin of the global frame to that point:

$$\mathbf{s}(a) = \mathbf{s}_0 + a\mathbf{u},$$

where the vector \mathbf{s}_0 points to a fixed position on the cluster, the unit vector \mathbf{u} points along the strip direction, and the variable a gives the distance from point s_0 to point $s(a)$. The track can be approximated in the vicinity of the cluster by a straight line, with any point $t(b)$ on that track given by the vector $\mathbf{t}(b)$:

$$\mathbf{t}(b) = \mathbf{t}_0 + b\mathbf{v}$$

where \mathbf{t}_0 points to a fixed position on the track, the unit vector \mathbf{v} points along the track, and the variable b gives the distance along the track from point t_0 to point $t(b)$.

The vector from cluster point $s(a)$ to track point $t(b)$ is

$$\begin{aligned}\mathbf{d}(a,b) &= \mathbf{t}(b) - \mathbf{s}(a) \\ &= \mathbf{d}_0 + b\mathbf{v} - a\mathbf{u}\end{aligned}$$

where $\mathbf{d}_0 = \mathbf{t}_0 - \mathbf{s}_0$, the vector between the fixed positions on the track and cluster, as shown in Figure 1. The 3D residual \mathbf{d} is found by minimizing length of $\mathbf{d}(a,b)$ with respect to a and b by requiring that $\partial|\mathbf{d}(a,b)|/\partial a = 0$ and $\partial|\mathbf{d}(a,b)|/\partial b = 0$, which gives

$$\begin{aligned}a_0 &= \frac{\mathbf{d}_0 \cdot (\mathbf{u} + \mathbf{v} \cos \gamma)}{1 - \cos^2 \gamma}, \\ b_0 &= \frac{\mathbf{d}_0 \cdot (\mathbf{v} + \mathbf{u} \cos \gamma)}{1 - \cos^2 \gamma},\end{aligned}$$

where $\cos \gamma = \mathbf{u} \cdot \mathbf{v}$ is the cosine of the angle between the track and cluster. The vector residual \mathbf{d} , representing the closest distance of approach, is then

$$\mathbf{d} = \mathbf{d}_0 + b_0\mathbf{v} - a_0\mathbf{u},$$

and is shown in Figure 1. The vector pointing to the origin of \mathbf{d} on the cluster is

$$\mathbf{s} = \mathbf{s}_0 + a_0\mathbf{u}$$

and the vector pointing to the terminus of \mathbf{d} on the track is

$$\mathbf{t} = \mathbf{t}_0 + b_0 \mathbf{v} .$$

Covariance matrix for residuals

The residual vector \mathbf{d} represents the displacement that, if applied to the silicon cluster, would align it with the track. Only one component of \mathbf{d} is well measured, the component along its length, with an uncertainty σ equal to the silicon strip resolution, which is of order ten microns. The components perpendicular to \mathbf{d} are unmeasured. In other words, the silicon wafer could be moved along the cluster or the track without changing \mathbf{d} .

To construct \mathbf{d} , imagine a vector \mathbf{x} which lies along the x-axis, has the same length as \mathbf{d} , and the same length uncertainty σ . It's covariance matrix would then be

$$\mathbf{E}_x = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \infty & 0 \\ 0 & 0 & \infty \end{bmatrix} \cong \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & 10^6 \sigma^2 & 0 \\ 0 & 0 & 10^6 \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 10^6 & 0 \\ 0 & 0 & 10^6 \end{bmatrix}$$

where the infinite errors have been approximated by $10^3 \sigma$.

Rotating this to the direction of \mathbf{d} gives

$$\mathbf{E} = \mathbf{R} \mathbf{E}_x \mathbf{R}^{-1} .$$

where \mathbf{R} is a rotation around the axis $\hat{\mathbf{x}} \times \mathbf{d}$ by an angle $\hat{\mathbf{x}} \cdot \hat{\mathbf{d}}$, so that $\mathbf{d} = \mathbf{R} \mathbf{x}$. This is the covariance \mathbf{E} of the 3D residual \mathbf{d} .

Combining 3D Residuals

Now consider forming a weighted average. For example, consider a track passing through a double-sided silicon wedge and perpendicular to it, as shown in Figure 2.

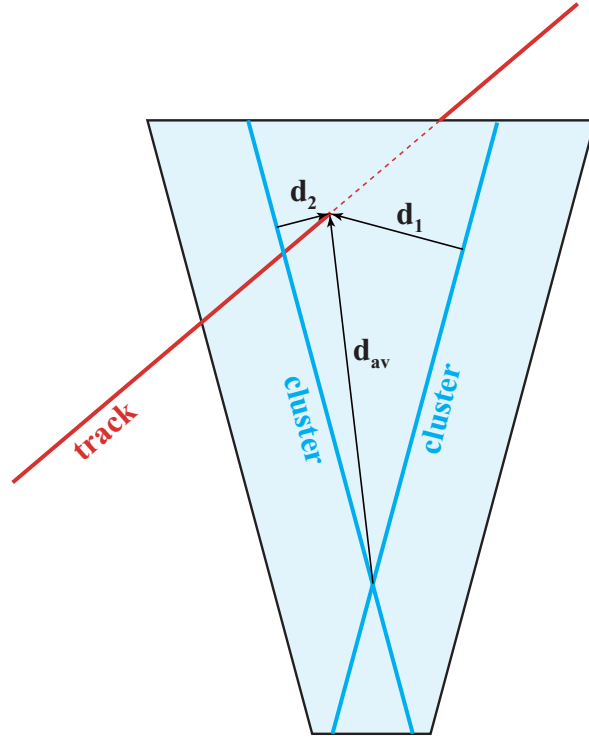


Figure 2. shows a two-sided silicon wedge with a track perpendicular to it and with two cluster measurements of the track. The weighted average of the two residuals gives the amount by which the wedge is misaligned.

There are now two residual vectors, \mathbf{d}_1 and \mathbf{d}_2 . The error matrix \mathbf{E}_1 represents an error ellipse that is finite only along the direction of \mathbf{d}_1 and infinite along the cluster and along the direction of the track. Likewise the error ellipse corresponding to \mathbf{E}_2 is perpendicular to the wedge and intersects it along the second cluster. The translation \mathbf{h} and its covariance matrix \mathbf{E} are given by the formulas for \mathbf{h} and \mathbf{E} above, for the special case when only translations are being considered:

$$\mathbf{h} = \left(\sum_i \mathbf{E}_i^{-1} \right)^{-1} \sum_i \mathbf{E}_i^{-1} \mathbf{d}_i, \quad \mathbf{E} = \left(\sum_i \mathbf{E}_i^{-1} \right)^{-1}.$$

The error ellipse corresponding to \mathbf{E} is inscribed in the diamond where the two clusters cross, and is infinite along the direction of the track, which is perpendicular to the wedge in this example. This means that the translation \mathbf{h} needed to align the detector is well measured in both dimensions in the plane of the silicon, but unmeasured along the track. This illustrates how two residual vectors, \mathbf{d}_1 and \mathbf{d}_2 , each well measured in only one dimension, combine to give an average residual, the translation \mathbf{h} , which is well measured in two dimensions. It also illustrates that the error matrices of the residuals carry information about the detector geometry, in this case, the strip orientations.

If several tracks strike the wedge at different angles, the weighted average of their residuals will have finite errors in all three dimensions. With at least three tracks at different angles, the translation and rotation alignment parameters of the wedge can be calculated. Additional tracks over-constrain its alignment parameters.

Appendix A. Proofs of Equations

Formulas for 3D residual vectors

Consider a rotation \mathbf{R} of the measurement positions \mathbf{s}_i followed by a translation \mathbf{h} to give new measurement positions \mathbf{s}'_i .

$$\mathbf{s}'_i = \mathbf{R}\mathbf{s}_i + \mathbf{h}$$

Fitted track positions \mathbf{t}_i remain stationary so that the original residuals \mathbf{d}_i are transformed to \mathbf{d}'_i as follows.

$$\begin{aligned} \mathbf{d}_i &= \mathbf{t}_i - \mathbf{s}_i \\ \mathbf{d}'_i &= \mathbf{t}_i - \mathbf{s}'_i \\ &= \mathbf{d}_i + \mathbf{s}_i - \mathbf{s}'_i \\ &= \mathbf{d}_i + \mathbf{s}_i - \mathbf{R}\mathbf{s}_i - \mathbf{h} \\ &= \mathbf{d}_i - (\mathbf{R} - \mathbf{1})\mathbf{s}_i - \mathbf{h} \\ &= \mathbf{d}_i - \mathbf{Q}\mathbf{s}_i - \mathbf{h} \end{aligned}$$

The \mathbf{d}'_i are linear in the values of the components h_i of \mathbf{h} since $\mathbf{h} = h_i \mathbf{u}_i$ where the \mathbf{u}_i are unit vectors along the axes. \mathbf{Q} can be linearized for small rotations θ_i around the x, y, and z axes using $\mathbf{Q} = \theta_i \mathbf{Q}_i$, with

$$\mathbf{Q}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \mathbf{Q}_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

This yields

$$\mathbf{d}'_i = \mathbf{d}_i - \sum_j \theta_j \mathbf{Q}_j \mathbf{s}_i - \sum_k h_k \mathbf{u}_k.$$

The $\chi^2(\mathbf{Q}, \mathbf{h})$ that \mathbf{d}'_i , the residuals after rotation \mathbf{R} ($\mathbf{Q} = \mathbf{R} - \mathbf{1}$) and translation \mathbf{h} , are equal to zero within their error matrices \mathbf{E}_i is

$$\chi^2(\mathbf{Q}, \mathbf{h}) = \sum_i \mathbf{d}'_i{}^\dagger \mathbf{E}_i^{-1} \mathbf{d}'_i,$$

$$\mathbf{d}'_i = \mathbf{d}_i - \mathbf{Q} \mathbf{s}_i - \mathbf{h}.$$

Differentiating with respect to each of the six parameters θ_j and h_k to minimize $\chi^2(\mathbf{Q}, \mathbf{h})$ yields

$$\frac{\partial \chi^2}{\partial \theta_j} = \sum_i (\mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}'_i + \mathbf{d}'_i{}^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_j \mathbf{s}_i), \quad j=1,3,$$

$$\frac{\partial \chi^2}{\partial h_j} = \sum_i (\mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}'_i + \mathbf{d}'_i{}^\dagger \mathbf{E}_i^{-1} \mathbf{u}_j), \quad j=1,3.$$

Because the error matrix is symmetric and real, $E_i^{-1} = E_i^{-1\dagger}$, so that the first terms of the above two equations are the transpose of the second terms. Since the terms are real scalar quantities, and since the transpose of a real scalar is equal to the scalar, the first and second terms are equal. The minimization condition that the six derivatives in the above equation are equal to zero then yields the six equations

$$\sum_i (\mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}'_i) = 0, \quad j=1,3,$$

$$\sum_i (\mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}'_i) = 0, \quad j=1,3.$$

or in terms of the six parameters θ_m and h_n ,

$$\sum_i \left(\mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \left(\mathbf{d}_i - \sum_m \theta_m \mathbf{Q}_m \mathbf{s}_i - \sum_n h_n \mathbf{u}_n \right) \right) = 0, \quad j=1,3,$$

$$\sum_i \left(\mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \left(\mathbf{d}_i - \sum_m \theta_m \mathbf{Q}_m \mathbf{s}_i - \sum_n h_n \mathbf{u}_n \right) \right) = 0, \quad j=1,3.$$

These six equations can be solved for the rotation and translation yielding minimum χ^2 . They can be written

$$\begin{aligned}\sum_{m=1}^3 A_{jm} \theta_m + \sum_{n=1}^3 B_{jn} h_n &= c_j, \quad j=1,3, \\ \sum_{m=1}^3 D_{jm} \theta_m + \sum_{n=1}^3 F_{jn} h_n &= g_j, \quad j=1,3,\end{aligned}$$

$$A_{jm} = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_m \mathbf{s}_i, \quad B_{jn} = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_n, \quad c_j = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i,$$

$$D_{jm} = \sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_m \mathbf{s}_i, \quad F_{jn} = \sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_n, \quad g_j = \sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i.$$

These six equations can be written as a single six-dimensional matrix equation

$$\mathbf{M} \mathbf{p} = \mathbf{k}$$

where the six-by-six matrix \mathbf{M} and six-vectors \mathbf{p} (parameters) and \mathbf{k} (constants) are given by

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{D} & \mathbf{F} \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{h} \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} \mathbf{c} \\ \mathbf{g} \end{bmatrix}.$$

where \mathbf{A} , \mathbf{B} , \mathbf{D} , and \mathbf{F} are 3x3 matrices and \mathbf{c} and \mathbf{g} are 3-vectors whose elements are given above. The minimum χ^2 solution for \mathbf{v} , the rotation-translation six-vector, is

$$\mathbf{p} = \mathbf{M}^{-1} \mathbf{k}.$$

To obtain the error matrix (or covariance matrix) \mathbf{V} of the translation and rotation parameters \mathbf{p} , we use

$$\mathbf{V}^{-1}_{jk} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial p_j \partial p_k}$$

Using the first derivatives from above, we get

$$\frac{1}{2} \frac{\partial^2 \chi^2}{\partial \theta_j \partial \theta_k} = \sum_i (\mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_k \mathbf{s}_i) = A_{jk}, \quad j,k=1,3,$$

$$\frac{1}{2} \frac{\partial^2 \chi^2}{\partial h_j \partial h_k} = \sum_i (\mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_k) = F_{jk}, \quad j,k=1,3.$$

$$\frac{1}{2} \frac{\partial^2 \chi^2}{\partial h_j \partial \theta_k} = \sum_i (\mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_k \mathbf{s}_i) = D_{jk}, \quad j,k=1,3,$$

$$\frac{1}{2} \frac{\partial^2 \chi^2}{\partial \theta_j \partial h_k} = \sum_i (\mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_k) = B_{jk}, \quad j,k=1,3.$$

This shows that $\mathbf{M} = \mathbf{V}^{-1}$, or that the solution can be written

$$\mathbf{p} = \mathbf{V} \mathbf{k}$$

which expresses the parameters \mathbf{p} in terms of their error matrix \mathbf{V} times the constant vector \mathbf{k} .

Special case: pure translation, no rotation

To understand these formulas, consider some special cases. If we set the rotation \mathbf{Q} to zero, then only F_{jn} and g_j are non-zero, yielding

$$\sum_{n=1}^3 \sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_n h_n = \sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i, \quad j=1,3.$$

or

$$\sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{h} = \sum_i \mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i, \quad j=1,3,$$

or

$$\left(\sum_i \mathbf{E}_i^{-1} \right) \mathbf{h} = \sum_i \mathbf{E}_i^{-1} \mathbf{d}_i,$$

or

$$\mathbf{h} = \left(\sum_i \mathbf{E}_i^{-1} \right)^{-1} \sum_i \mathbf{E}_i^{-1} \mathbf{d}_i,$$

The error matrix for the vector \mathbf{h} is

$$\mathbf{E}^{-1}_{jk} = \frac{1}{2} \left(\frac{\partial^2 \chi^2}{\partial h_j \partial h_k} \right) = \sum_i (\mathbf{u}_j^\dagger \mathbf{E}_i^{-1} \mathbf{u}_k)$$

$$\mathbf{E} = \left(\sum_i \mathbf{E}_i^{-1} \right)^{-1}$$

In other words, for purely translational alignment, \mathbf{h} is the weighted average of the residuals, weighted by the inverses of their error matrices. The error matrix \mathbf{E} for \mathbf{h} is just the inverse of the sum of the weights.

Special case: pure rotation, no translation

Another special case is that of purely rotational alignment so that translation \mathbf{h} is zero so that only A_{jm} and c_j enter so that

$$\sum_{m=1}^3 A_{jm} \theta_m = c_j, \quad j=1,3,$$

$$A_{jm} = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_m \mathbf{s}_i, \quad c_j = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i$$

$$\theta_m = \left(\sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{Q}_m \mathbf{s}_i \right)^{-1} \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i$$

where the expression in parentheses is supposed to mean the matrix A_{jm} or, in more clear matrix-vector notation,

$$\mathbf{A} \boldsymbol{\theta} = \mathbf{c},$$

$$\boldsymbol{\theta} = \mathbf{A}^{-1} \mathbf{c},$$

The error matrix for the vector $\boldsymbol{\theta}$ is \mathbf{A}^{-1} where the elements of \mathbf{A} are given above. The rotation vector $\boldsymbol{\theta}$ is, as was seen above for translation \mathbf{h} , a weighted average over the residuals \mathbf{d}_i , but with a more complex weighting. Note that $\mathbf{Q}_j \mathbf{s}_i$ is the cross product of a unit vector in the \mathbf{j}^{th} direction with the measurement location vector \mathbf{s}_i , so it is a vector in the direction that the measurement would move for a rotation about the \mathbf{j}^{th} axis. This means that the weighting takes into account that for a rotation, each residual moves in a different direction, depending on its measurement location \mathbf{s}_i .

Another way to get a feeling for the meaning of this special case is to using vector analysis notation, in which

$$\sum_m \theta_m \mathbf{Q}_m \mathbf{s}_i = \boldsymbol{\theta} \times \mathbf{s}_i,$$

where $\boldsymbol{\theta}$ is the axial vector with components θ_m , so that

$$\sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} (\boldsymbol{\theta} \times \mathbf{s}_i) = \sum_i \mathbf{s}_i^\dagger \mathbf{Q}_j^\dagger \mathbf{E}_i^{-1} \mathbf{d}_i$$

from which one can see that if the \mathbf{d}_i are just displacements due to the small rotation $\boldsymbol{\theta}$ with no other source of residuals, then $\mathbf{d}_i = \boldsymbol{\theta} \times \mathbf{s}_i$ and the equation is exactly satisfied.